Comparative Analysis of Infrastructure Assessment Methodologies at the Small Unit Level

Marc J. Distefano[1] and William J. O'Brien, Ph.D., M.ASCE[2]

Abstract: This paper presents a comparative analysis of three different infrastructure assessment tools currently used by the U.S. Army's combat engineer small units (squad, platoon, company). The infrastructure assessment tools included one pen and paper checklist and two software-based tools based on specialized handheld hardware for data collection in combat conditions. Evaluation was conducted using macroergonomic and applied cognitive task analysis methods with soldiers of varying levels of infrastructure assessment expertise. Each assessment method was evaluated based on performance (time, errors, and accuracy) and usability (ergonomic and cognitive challenges). We found that soldiers documented more accurate data using handheld digital devices instead of pen-and-paper assessment forms, but there was little difference in time between the methods. Contrary to initial expectation, we also found that most soldiers prefer a slightly more difficult to use handheld assessment device with customized checklists rather than an easier to use, less restrictive device. Critiques developed by this research can be used to improve engineering and construction organizations as well as add to our design guidelines for portable infrastructure assessment tools and methods.

DOI: 10.1061/(ASCE)0733-9364(2009)135:2(96)

CE Database subject headings: Infrastructure; Human factors; Computer analysis; Assessments.

Introduction

The infrastructure assessment process supports the brigade combat team's (BCT's) assured mobility mission, which is critical for stability operations in theaters such as Iraq. This process involves several organizations collecting, organizing, and processing infrastructure data to support the combatant commander's effects-based operations. However, this process is not as efficient as it can be primarily given the constraints combat places on time, performance, quality, security, and availability of personnel. In most cases, the personnel with the most infrastructure experience cannot easily conduct assessments based on the threat level. Rather, it is the small unit leaders who are the most available and properly resourced for combat operations that conduct assessment. Typically, these small unit leaders lack the expertise to conduct an assessment without assistance or specific instructions.

The Army is currently developing and/or fielding assessment tools to bridge this gap between experience and availability. However, whatever tool is used at the BCT level, there are certain conditions that should be met (Andrysiak, personal communication, 2006). Any small unit infrastructure assessment tool at the tactical level must be: easy enough for small unit leaders with limited infrastructure knowledge to learn and use; durable and small enough for a soldier to comfortably use in the field; effective enough to capture critical data in fewer missions to reduce the prolonged exposure of troops on the ground; and be able to facilitate rapid data collection, which could be efficiently processed, cataloged, and managed at the brigade engineer level.

The primary objective of this research is to compare some of the existing small unit infrastructure assessment tools from a human factors approach to identify the cognitive and organizational challenges that affect performance and usability. While conducted in the context of a military application, the challenges and requirements for portable infrastructure assessment tools that can be deployed by users with limited experience are generalizable to a variety of engineering and construction applications, whether they are rapid response to disasters or inspections reporting back to centralized engineering resources.

Literature Review

Our research was motivated by this pressing issue: given the recent reorganization of the BCT and the limited operational time a unit has in Iraq, how can small units identify reconstruction projects in their area of operation more effectively and in less time? There are tools that exist that attempt to address this issue, but they have only been evaluated to a limited extent. As such, what are the cognitive, behavioral, and organizational constraints that prevent these tools from being more effective? To answer these questions, we reviewed previous research in small unit infrastructure assessments, usability evaluations of handheld devices from an individual and organizational perspective, human computer interaction (HCI), and cognitive task analysis techniques used in organizations with rapid product fielding requirements.

Infrastructure Assessment Research

Starting in 2003, cadets and faculty members in the department of civil and mechanical engineering at the United States Military

[1]Major, U.S. Army, Dept. of Systems Engineering, United States Military Academy, West Point, NY 10996.

[2]Assistant Professor, Construction Engineering and Project Management Program, Dept. of Civil, Architectural, and Environmental Engineering, Univ. of Texas at Austin, One University Station C-1752, Austin, TX 78712-0273 (corresponding author).

Note. Discussion open until July 1, 2009. Separate discussions must be submitted for individual papers. The manuscript for this paper was submitted for review and possible publication on May 3, 2007; approved on February 26, 2008. This paper is part of the ***Journal of Construction Engineering and Management***, Vol. 135, No. 2, February 1, 2009.

Report Documentation Page			*Form Approved OMB No. 0704-0188*
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.			
1. REPORT DATE **2009**	2. REPORT TYPE		3. DATES COVERED **00-00-2009 to 00-00-2009**
4. TITLE AND SUBTITLE **Comparative Analysis of Infrastructure Assessment Methodologies at the Small Unit Level**			5a. CONTRACT NUMBER
			5b. GRANT NUMBER
			5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)			5d. PROJECT NUMBER
			5e. TASK NUMBER
			5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Major, U.S. Army, Dept. of Systems Engineering,United States Military Academy,West Point,NY,10996**			8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSOR/MONITOR'S ACRONYM(S)
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**			
13. SUPPLEMENTARY NOTES			
14. ABSTRACT			
15. SUBJECT TERMS			

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**13**	

Academy researched effective infrastructure assessment programs and methodologies at the small unit level (Welch et al., unpublished report 2006). The result of 3 years of research and testing was a system of assessment sheets for five infrastructure types: power production, water treatment, transportation, waste water treatment, and trash disposal. With the assistance of the National Training Center, the US Army Engineer School, and the US Army Construction Engineering Research Laboratory (CERL), the cadets tested soldiers' abilities to fill out basic questionnaires supplemented with photos and descriptions of infrastructure. This research led to the creation of the sewage, water, electrical, academic, and transportation (SWEAT) assessment forms, which were modified to cargo pocket size for easy storage and accessibility in the field. In addition to developing the first SWEAT manual, the cadets found the following: (1) the assessments should be done completely in electronic form to include collecting data on a personal digital assistant (PDA) that included a digital camera, global positioning system (GPS), and a digital depth/length finder; (2) the assessments should have a score that can be directly used to help commanders and staff officers decide on how much resources should be allocated to a particular facility or components of a facility in a particular area of operations; and (3) all of the assessments should be included in a large scale infrastructure overlay for commanders at every echelon to access and review.

The U.S. Army Construction Engineering Research Laboratory conducted an evaluation and technical report describing the effectiveness and usability of the handheld apparatus for mobile mapping and expediting reporting (HAMMER) in 2006 (Britt 2006). The report consisted of an after-action review of a large scale assessment project at Fort Jackson, South Carolina, which was to locate electrical, natural gas, water, and wastewater features on the military installation using submeter global positioning system survey. The participants in this study were civilian employees familiar with HAMMER technology and were not constrained to time. During this evaluation, which consisted of documenting over 4,000 GPS points and data entries covering 4,000 acres of cantonment area, CERL documented the results and findings of the evaluation to include human factor issues that affected performance and usability. The key benefits to using the HAMMER were: (1) the interface between the handheld PDA and other devices was easy to use; (2) the laser distance meter provided a means of capturing GPS data for features not easily accessed; (3) digital photos were extremely beneficial during data postprocessing; (4) the HAMMER was relatively light (2 lb) and was not cumbersome in the field even after using it for a few hours; and (5) the stylus was attached to the device, which helped to prevent accidentally losing it in the field. CERL also documented the main ergonomic issues of the evaluation: (1) the digital cameral lens fogs, making it impossible to acquire a feature photo in damp weather conditions; (2) occasionally, the user interface "freezes" due to human error such as moving while taking images, which stops data collection until a soft reset is performed; and (3) the short battery life required users to carry a bulky external battery pack. Last, CERL made the following recommendations for the handheld assessment device: (1) refine the postprocessing to make it more user friendly and easier to understand; (2) simplify the upload and download process; (3) simplify the external battery connections and cabling.

The CERL studies are amplified by Pena-Mora and Dwivedi (2002), who argue that the demands of modern large-scale projects require a collaborative system that enables access to information and applications from anywhere with minimum device specifications. Their research centered on the comparison of usability and performance of multiple devices to include networking capable PDAs that project managers could easily use in the field. They found that a collaborative project management system with a knowledge repository, analysis resources, and multiple device access to include PDAs, PCs, and the internet, was instrumental in supporting the infrastructure of geographically distributed project management teams in complex projects (Pena-Mora and Dwivedi 2002). Furthermore, they found that the effectiveness of such a system is incumbent upon: (1) the project information must be able to be shared with any of the project management personnel, and (2) there had to be few limitations to the computing device (Pena-Mora and Dwivedi 2002).

Usability and Performance of Handheld Devices

There is alot of literature on the usability of handheld devices. We limit our review to literature specifically related to performance and usability comparison between traditional methods and handheld devices for discrete tasks. For example, Segall et al. (2004) compared the effectiveness of paper-and-pen type of checklists versus handheld methodologies by conducting an evaluation consisting of 34 college-aged students taking a PDA-based quiz and paper-and-pencil-based quiz. Their objective was to evaluate the PDA's usability in terms of effectiveness, efficiency, and user satisfaction (Segall et al. 2004). Through a series of statistical tests and user satisfaction surveys, the authors found that: (1) quiz scores of the students using both methods were not significantly different than each other; (2) the time spent on the quizzes was significantly lower for the PDA-based quiz; (3) the satisfaction of the two quiz types were not significantly different; and (4) no relationship existed between computer anxiety and satisfaction for either quiz types (Segall et al. 2004). We note that the authors did not closely examine the cognitive challenges that may or may not have impacted the quantitative results.

Connelly et al. (2005) evaluated usability of handheld devices to create an interface for dialysis patients of varying visual acuity, literary skills, and computer experience (Connelly et al. 2005). Their work was driven by the fact that patients who typically rely on manually recording and monitoring their diet had a very low rate of compliance where those who used electronic devices, such as PDAs, had a much higher rate of compliance (Connelly et al. 2005). Of the 30 patients of mixed demographics who volunteered for the study, they found that healthy older patients could physically interact with PDAs as well as healthy young patients (Connelly et al. 2005). Interaction was defined as pressing buttons, recording messages, and scanning bar codes. However, the authors did find that most of the participants had difficulty using the stylus to enter data and viewing the small icons and font on the screen. Their findings led them to design a system that took into consideration improved input methods and interface layouts (Connelly et al. 2005).

Of the many attributes that have been consistently recorded on the usability of the PDAs such as their convenient size, light weight, and reasonable computer power, the size of the screen is one attribute that is commonly referred to as problematic. Karkkainen and Laarni (2002) evaluated this feature in depth. They highlighted their findings as follows: (1) users spend much more time manually scrolling or paging the text on a PDA due to the limited amount of information that can be displayed on a screen; (2) poorer display resolution of PDAs may lead to poorer performance in information retrieval tasks; and (3) virtual keyboards and/or handwriting recognition software often reduces screen

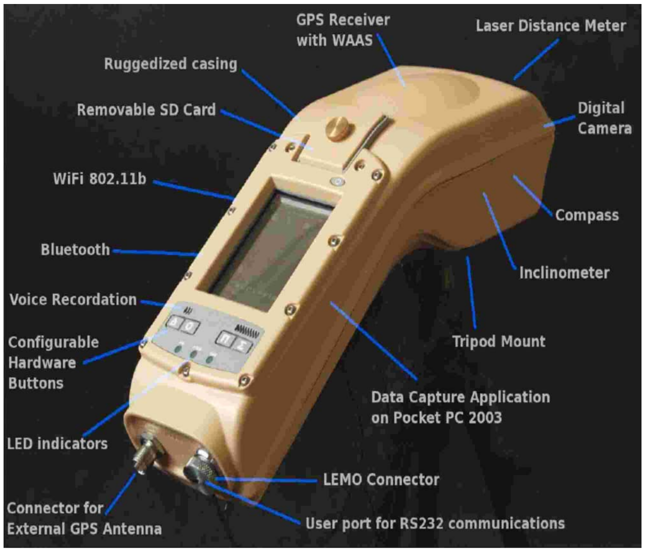


Fig. 1. IKE 304 handheld device platform

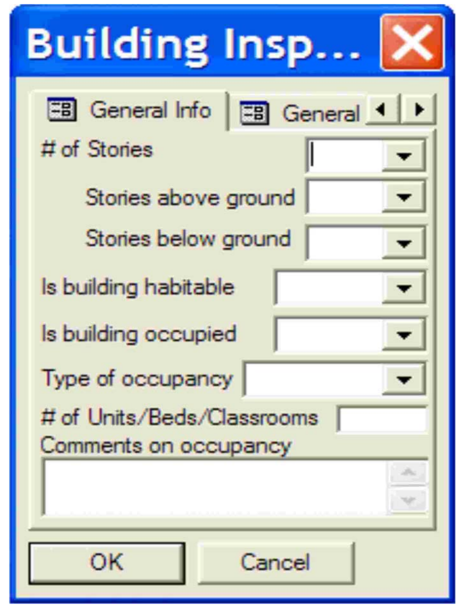


Fig. 2. HAMMER customized checklist screen shot

space even more and can be inaccurate (Karkkainen and Laarni 2002). They recommend designers consider these constraints when developing applications for PDAs.

Subject of Study

The three small unit level assessment tools we evaluated were: (1) the Army's sewage, water, electric, academic (schools and related buildings), trash infrastructure reconnaissance (SWEAT-IR) manual, (2) the geospatial assessment tool for engineering reachback (GATER), and (3) the handheld apparatus for mobile mapping and expediting reporting (HAMMER) with customized assessment checklists. Both the GATER and HAMMER are available on the IKE-304 handheld digital device, which integrates a PDA, a GPS receiver, a laser range finder, a digital compass, and a digital camera as seen in Fig. 1.

SWEAT-IR Manual

The *SWEAT-IR Manual* is designed to broadly describe the infrastructure reconnaissance process, including the relevance of assessments at the small unit level, whose military occupational specialties are included in the process, how typical infrastructure facilities operate, and what basic components of each facility type look like and their functions. After each section of the manual, which corresponds to an infrastructure subset (i.e., SWEAT), soldiers can use easy to understand checklists to document critical data for subsequent military decision making at the BCT staff level. These checklists are referenced with recognizable icons that help guide the user to take particular actions during the checklist such as take a digital image, take a measurement, "eyeball" a measurement, and/or record the geographic coordinates of the site. As such, it is important to note that the SWEAT-IR manual is not entirely a stand alone tool, and is recommended for use with supplemental tools such as a camera.

HAMMER

The HAMMER is a remote data capturing device and application that has a battery life of 8 hours, takes 3 hours to fully charge, operates on a 624 MHz Intel PXA 270 processor, has 128 MB of RAM, 1 GB storage data card, a $3'' \times 2''$ touch screen with virtual keyboard. The HAMMER integrates the IKE 304 hardware with ESRI ArcPad Application Builder Software, which allows users to download GIS-based map data from ArcMap, download the data to the device, and access, capture, and edit data on the device via ArcPad. This software allows users to document data in a database format where each assessment is a record and individual attributes from the assessment are saved as fields within that record. Mission planners use the software to create checklist-like forms that can be downloaded to the device as layers, which helps to direct the field data collector to capture specific information, which facilitates organized data capture, transfer, and eventual postprocessing as seen in Fig. 2. As such, there is considerable preplanning for optimal use of the HAMMER. The HAMMER also requires users to follow a rigid checklist procedure to record data.

GATER

The GATER utilizes the same hardware as HAMMER. However, there are some distinct differences between the applications. First, the GATER application is a web-based tool that allows users to download map data from a secured website to the device, collect data on standardized forms, download the data back to a PC, view summarized reports, and submit the data back to the system's proponent for further data processing and collaboration. Second, data collection with GATER is generally a process of taking images with geospatial attributes and an open-ended description that the user inputs. There are few checklists for users to follow during the data collection process. Furthermore, the application is designed to be available to units regardless of military occupational specialty and does not require that the military planner or the soldiers collecting the data be trained in ESRI ArcPad Application Builder Software. As long as the user has access to the GATER's secured website, has a network capable PC, and the GATER device itself, they can use it anywhere and share information among many authorized users. Fig. 3 represents the basic GATER screen shots, which represent capturing and documenting data.

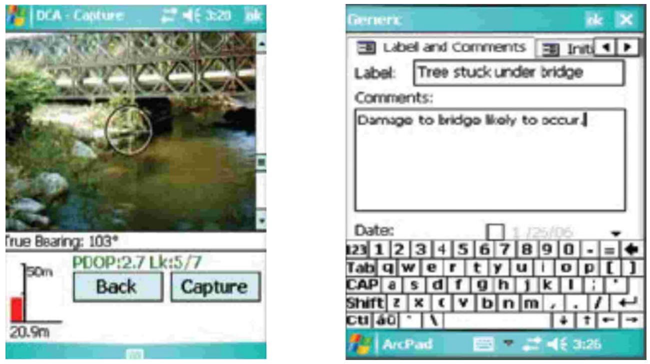


Fig. 3. GATER data collection screen shot

Research Methodology

Our research methodology consisted of identifying research variables to measure, identifying those hypotheses that would form the basis of our research, identifying the evaluative techniques that would best structure our research, and evaluating participants that would most likely use the assessment tools in a field experiment.

Research Variables

We defined performance in terms of accuracy and time. Both accuracy and time were compared against a "standard" assessment that we created with subject matter experts prior to the experiment. We defined "accuracy" as how many correct data entries (descriptions and images) a soldier collected during an assessment. We defined "time" as how long it took a soldier to complete an assessment from start to finish.

The usability evaluation was much more qualitative than the performance evaluation. We defined usability as how difficult it was for soldiers to accomplish the tasks associated with the specific tool type in terms of cognition and interaction with the device. We further defined usability in terms of how well a soldier would perform given the organizational issues related to his/her job.

Research Hypotheses

We developed three hypotheses about the performance and usability of small unit infrastructure assessments.

Hypothesis 1. Small unit leaders will have similar performance scores across all three assessment methodologies.

Hypothesis 2. Small unit leaders should complete infrastructure assessments using handheld devices faster than the paper-and-pen checklists.

Hypothesis 3. Small unit leaders should find a more user friendly and less technical assessment methodology more desirable than a more difficult to use methodology.

These hypotheses are confirmed by the literature review indicating that performance on tests is similar for pen and paper and PDAs, relative increases in speed with electronic devices, and user preferences for easier to use tools.

Research Techniques

To explore these hypotheses, we needed to evaluate the existing tools for this research from a structured approach that addresses the cognitive and ergonomic issues that affect performance and usability. Evaluative methods used were applied cognitive task analysis (ACTA) and macroergonomics, which we used in tandem to identify and describe those issues that either promote or prevent efficient infrastructure assessments at the small unit level. Fig. 4 represents a basic flowchart of how these techniques were used.

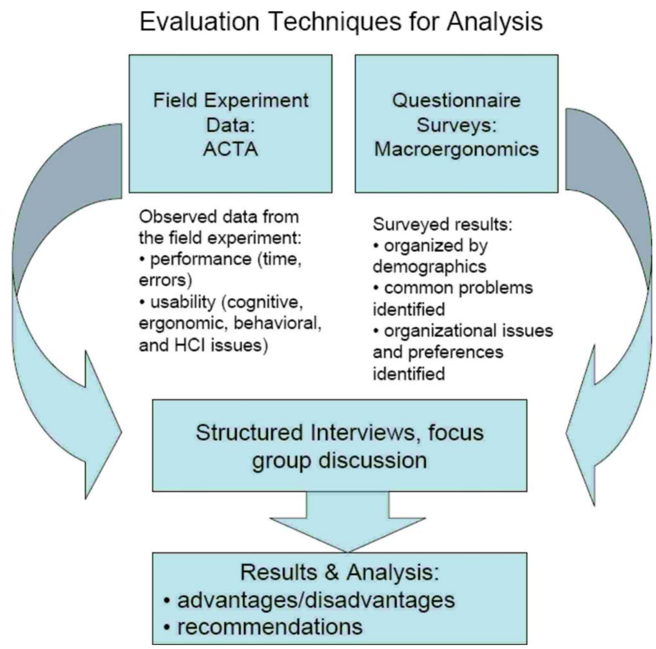


Fig. 4. Research methodology flowchart

Applied Cognitive Task Analysis

We employed ACTA techniques to elicit critical information from tool developers and senior engineer officers who are very familiar with infrastructure assessments and scope of work development. The ACTA techniques begin with interview methods that build on the subject matter expert's infrastructure assessment skills to include diagnosing, predicting, situational awareness, improvising, metacognition, and compensating for equipment limitations (Militello and Hutton 1998). These interview techniques help to structure the analysis in which evaluators can monitor soldiers during their use of the tools and document findings on a cognitive demands table. This table helps to list what the cognitive elements of each analysis task were, why they were difficult, what the common errors might be, and cues and strategies that could be observed (Militello and Hutton 1998). Moreover, the table's purpose is to provide a format to use in focusing the analysis of the project goals: to elicit those cognitive issues small unit leaders might encounter when using the tools.

Macroergonomics

The third phase of the research consisted of providing a macroergonomic organizational questionnaire survey (MOQS) used to assess individual and organizational performance and usability. The MOQS is a useful tool that can be used throughout the design, development, testing, and implementation phase of a product or service (Carayon and Smith 2000). To develop the survey, we first submitted an online survey to senior faculty and engineer officers at West Point to better test the relevance and meaningful-

Table 1. Accuracy Scores per Tool Type (Accuracy in %)

	SWEAT		HAMMER		GATER	
Tool group	1BCT (%)	1-395 (%)	1BCT (%)	1-395 (%)	1BCT (%)	1-395 (%)
Individual soldier scores	51.72	72.41	58.62	86.21	51.72	62.07
	41.38	70.69	51.72	72.41	55.17	58.62
	51.72	75.86	51.72	82.76	58.62	72.41
	51.72		68.97		68.97	
			86.21		72.41	
			79.31			
Group average	Average 49.14	Average 72.99	Average 57.76	Average 80.46	Average 58.62	Average 64.37
Total average	Total average 59.36		Total average 70.88		Total average 62.50	

ness of the questions on the survey (Solomon 2001). Substantive questions were also designed to avoid bias in the form of question and rating scale (Converse and Presser 1986). The survey was separated into two areas. The first area asked participants to state whether or not they agreed with statements about the general usability of the tool themselves. The second part of the survey consisted of similar styled statements, but they were designed to get feedback on whether or not participants felt certain usability related issues were important to their job or to their organization.

Last, we conducted a macroergonomic focus group after the survey to get feedback on the tools to see how best to integrate the tools within the BCT (Newman 2005). The objective for the macroergonomic focus group was to expound on three areas that affect performance and usability at the soldier level both individually and at the organizational level. Those areas were: (1) training; (2) data collection; and (3) organizational structure.

Participants and Procedures

Two groups were studied. First, we evaluated performance with 12 soldiers from the 1st BCT, 1st cavalry division prior to their deployment to Iraq in Sept. 2006. The second group consisted of 15 soldiers from the 1-395 combat engineer battalion (reserve). Of the soldiers evaluated, most were small unit leaders (sergeant to sergeant first class within the enlisted ranks, second lieutenant to captain within the officer ranks). Ages ranged from 19 to 45 years old. Almost two-thirds of the soldiers were from combat engineer battalions.

Field experiments took place at Fort Hood with a simulated assessment developed with engineering personnel on base. During the field experiment, soldiers were separated into groups representing a particular assessment tool. The integrity of each group remained the same throughout the experiment. After receiving one day of basic instruction on infrastructure systems to include a walk through of some facilities, soldiers were divided into subgroups of tool type. In round robin format, soldiers conducted the assessment after receiving *task* (document as many faults that might exist), *conditions* (try to complete within 25 minutes; enemy contact likely), and *standards* (equipment and gear required). While the soldiers conducted the assessment, two designated observers recorded time and performance related errors. Additionally, a moderator captured cognitive issues and recorded them on a cognitive demands table for further analysis. Upon completion of the assessment, soldiers then participated in an organizational survey and subsequent focus group to expound on some of the usability issues that they had or could perceive to have given the conditions of combat and BCT organizational constraints.

Results and Findings

Accuracy Performance Comparisons

Our initial hypothesis is that we expected to find that small unit leaders had similar results for accuracy across each assessment method. We defined accuracy as the number of correct data entries filled out compared to the “gold standard” evaluation. Table 1 is a summary of accuracy scores per tool among the two groups of small unit leaders from 1st BCT and the 1-395 EN BN (reserve).

Our findings suggested that our data contradict our hypothesis because we can see that the scores are not similar between the three tools. Due to a low *N* value for each tool type, we did not prove statistical significance among these differences. However, we conclude from the high degree of difference between scores that most soldiers performed better with the handheld devices than with the paper-and-pen SWEAT checklists. That is, soldiers using these tools had the least number of errors observed during the experiment (entries skipped, entries entered incorrectly, actions not taken when required such as measure, capture image, or take grid coordinate). There are many reasons for this: (1) the HAMMER and GATER could take most of the geospatial data in a fraction of the time it took soldiers using the manual forms with a separate digital camera and handheld GPS with little human error. Those with the SWEAT manual had to document data separately (take image, take GPS grid, record data); and (2) soldiers using HAMMER and GATER could make entries while moving around the site unlike the soldiers using the manual forms who had to stop and write in order to be legible. Those using the SWEAT manual had to hand write necessary detail, and poorly written entries often had to be erased and rewritten for clarity. This was especially difficult in adverse weather conditions such as high wind.

We did see from Table 1 that there was a high degree of deviations in the maximum and minimum scores among tool types,

Table 2. Accuracy Scores by Age Group

Tool used	Accuracy (in %) 20–25 years	26–29 years	30–35 years	36+ years
HAMMER	57	82	72	71
GATER	43	60	62	62
SWEAT	42	52	62	73

so we further organized the data into age groups that closely represent different levels of small unit leader in Table 2. The 20–25-year old age group closely represents team leaders and platoon leaders that have little real world combat experience. The 26–29-year old age group closely represents senior lieutenants, junior captains, and squad leaders. The 30–35-year olds represent more senior captains and platoon sergeants. Finally, the 36+-year olds closely represent those officers and noncommissioned officers beyond company level.

We found that the HAMMER had higher accuracy scores among most age groups, especially within the 26–29-year olds, which is more representative of a small unit leader (squad leader, platoon leader). The HAMMER also performed better than the GATER in each age group. The reasons for this are because: (1) the HAMMER has dropdown menus for certain fields that did not require the soldier to type any information, unlike the GATER where most of the data collected must be documented using the stylus and virtual keyboard; (2) the HAMMER has entries that require specific information that most closely resembles that of SWEAT checklists. Other than one entry indicating the overall condition of a facility, the GATER's application does not require a user to fill in specific infrastructure data to be "complete."

A second important finding from Table 2 is that experience matters with regard to level of accuracy and that familiarity with the tools alone does not promote higher scores. There are several reasons for this as well: (1) the older a soldier, the more likely it is he/she has more infrastructure knowledge and/or experience on infrastructure assessment in combat. For example, the 20–25-year old age group who is probably more familiar with new technology (cell phones, PDAs, compact video games) scored less than 60% with all tools, which suggests that infrastructure knowledge was their limiting factor. On the contrary, the age groups representing 30+ years had average scores with all tools above 60%.

The only exception to the finding that accuracy improves with age is with the HAMMER in the 26–29-year old age group, which decreases after 30+ years. We conclude that within this group, there are several soldiers who have been involved with infrastructure assessments in Iraq recently and understand the basic assessment process. They also are more familiar with some of the Army's new digital technology such as military grade GPS receivers and transmitters, laser range finders, and the force XX1 battle command brigade-and-below command and control network system. Many within this age group stated that their experience troubleshooting these types of systems prepared them to troubleshoot some of the HAMMER's issues.

Time Performance Comparison among Methodologies

Our second hypothesis was that most soldiers would perform faster using handheld devices than with the SWEAT checklists. We define time as the duration of the assessment from start to finish. We established that 25 min was the standard to better represent realistic assessment times in combat. Any longer than 25 min could significantly leave data collectors exposed during real world assessment missions in combat.

Our findings contradict our hypothesis that data collectors using handheld devices would be significantly faster than the SWEAT checklists as seen in Table 3.

We can conclude from Table 3 that most of the average times per tool are within 6 min of the 25 min standard without a significant deviation between the average times. Again, our low N values do not support a statistically significant finding that tool type contributes to a faster or slower time. However, we can conclude from the small deviation between the two units average time using HAMMER that there is a low degree of difference among soldiers who used the HAMMER. They were consistently faster than or almost as fast as the standard. From our observations, possible reasons for this are: (1) the HAMMER provides users with structured checklists that help guide soldiers through the assessment process; (2) data collection with the assistance of dropdown menus and multiple choice options that can be selected with the tap of a stylus is much faster than hand writing comments on either paper or the GATER's virtual keyboard; and (3) users spend less time reviewing saved data with HAMMER because they can scroll through the entry "pages"quickly.

Table 3. Time Scores per Tool Type (Time in Min)

Tool	SWEAT		HAMMER		GATER	
group	1-BCT	1-395	1-BCT	1-395	1-BCT	1-395
Individual soldier scores	17.00	41.00	21.00	23.00	27.00	38.00
	22.00	32.00	19.00	21.00	32.00	24.00
	21.00	29.00	27.00	27.00	34.00	33.00
	29.00		35.00		29.00	
			15.00		28.00	
			24.00			
Group average	Average 22.25	Average 34.00	Average 25.50	Average 23.67	Average 30.50	Average 31.67
Total average	Total average 27.29		Total average 24.71		Total average 31.00	

Table 4. Time Scores by Age Group

Tool used	Time (in min) 20–25 years mean	26–29 years mean	30–35 years	36+ years
HAMMER	24.50	22.80	21.00	29.67
GATER	25.86	25.00	34.00	32.67
SWEAT	24.67	21.67	32.50	34.50

For further insight into the time differences, we organized the data into age groups as seen in Table 4.

From Table 4, we conclude that most of the average times per tool are within 6 min of the standard without significant deviations between each other. There is a low degree of difference between the 20–29-year old ranges, but there is a high degree of difference among the 30+year range.

The 20–25-year old age group, on average, lacks infrastructure assessment knowledge and experience; therefore, most did not document much detail with any of the tools. For the most part, GATER and SWEAT do not have any designed restrictions within their process that prevents users from skipping entries or filling in incomplete data in certain cases. The HAMMER has dropdown menus and multiple choice features, which encourages users to answer, but the system has few required fields. We could not conclude that the 26–29-year old group performed faster with handheld devices either. This group performed slowest with the GATER. We attribute this to the fact that most of this group had recent experience in Iraq where many did conduct infrastructure assessments of some type and felt like a checklist, either digital or manual, was useful in collecting data, not necessarily the unstructured format of GATER.

The older two age groups had a higher degree of difference in time among the different methodologies. However, the differences do not support that they are categorically faster using handheld devices. For example, the 30–35-year old age group performed fastest with the HAMMER but performed slowest with the GATER. The 36+year olds also performed fastest with the HAMMER but performed slowest with the SWEAT checklists. A possible explanation for this could be that the older age groups have more experience with more complex reporting technologies used during Operation Iraqi Freedom.

Soldier's Preference of Easier to Use Handheld Device

Our third hypothesis was that small unit leaders would prefer a handheld device based on fewer technical functions, simplicity of use, and the multiple uses the devices can perform based on lack of rigid checklists. To compare usability in terms of satisfaction of the ergonomics of the tools, how challenging tasks and troubleshooting may have been, and/or how effective the soldiers may perceive the tool to be within the context of their military occupational specialty, we employed macroergonomic techniques to derive our usability results and findings. Our findings suggested that while most soldiers found the GATER to be easier to use, most found that the BCT would be better served with a tool that is more difficult to use but facilitates more detailed data collection.

Of the 24 soldiers who conducted the evaluation, 19 filled out a survey. As mentioned earlier, the survey was organized into two areas: (1) whether or not participants agreed or disagreed with certain usability statements regarding a tool type; and (2) whether or not participants felt certain features were important or not important to their jobs and/or organization. As each survey question had a numerical value range (1 to 5), we obtained the average score per question of all surveyed participants. Table 5 represents the questions that we felt were the most revealing on the overall usability of the tools themselves and how important the features of the tools were to the soldiers' jobs and to their organization.

From the survey results, we conclude that most soldiers felt that handheld devices would be better for documenting data in real world situations where they might have to react to contact. Most soldiers also felt that size of the device and format structure was important for their jobs while stating that an easy to use tool is not necessarily important. In addition to surveys, we also asked soldiers to include remarks for those questions that they strongly agreed/disagreed and/or felt was very important/not important. Of the many individual comments, we felt that the remarks listed in Table 6 are very representative of the overall sentiment towards each tool type.

To better explain some of the main usability comments from the surveys, we organized the following topics that we discussed in further detail during the macroergonomic focus group: (1) react to contact (simulated during the field experiments); (2) recommend a device to chain of command; (3) whether soldiers could

Table 5. Summarized Survey Results of Most Revealing Questions

Strongly agree (1=strongly disagree, 5=strongly agree)	SWEAT (n=6)	GATER (n=6)	HAMMER (n=7)
React to contact while conducting an assessment	1.50	4.17	3.75
Recommend this to my commander as a primary assessment device for the unit	2.67	3.33	3.79
Use this in a real world assessment	3.92	4.33	4.38
Document more data using this tool than what I might have done without it	3.92	3.50	4.75
Document details in the given space	3.83	4.00	4.08
Very important (1=very important, 5=not very important)	**SWEAT (n=6)**	**GATER (n=6)**	**HAMMER (n=7)**
A tool that is more difficult to use but has a checklist as a guide	1.83	2.67	1.42
The size of the device	1.92	2.50	1.25
A tool that is easy to use but does not provide the most detailed information	3.83	2.83	4.08
A tool that can be used for other reporting purposes	1.92	2.00	1.63

Table 6. Soldier Remarks after Survey

Tool	Survey remarks
SWEAT	The space available on the checklist is inadequate, especially if condensed to pocket size.
	It is difficult to use with gloves and eye protection.
	It would be difficult to have a camera in one hand, checklist in the other, and have to react to enemy contact.
	The biggest limitation to the handheld digital camera is that I cannot review the images as they relate to a description.
	I feel like I am jumping around the checklist because it does not flow from top to bottom.
HAMMER	I am not confident that this tool will really help me with my job.
	I see little value in a system that cannot be shared by everyone.
	It is hard to see the screen in direct sunlight.
	If I hit the wrong button or command, I have to start over in most cases.
	I need much more training on this device.
	I could not take as many pictures as I needed with the digital checklist.
GATER	I like the GATER because it does not limit or impede my moment on the ground.
	The loading of the system takes too long.
	I am not confident that the training I received was adequate.
	I could not troubleshoot the GATER and did not understand how to refer to my assessment after completion.
	The virtual keyboard is not easy to use because it blocks most of the screen and I have to tap multiple times in some cases.

document more using a particular tool; and (4) whether a tool that is easier to use is more desirable than a more difficult tool that yields more information.

React to Contact

Most soldiers disagreed that it was easy to maneuver and react to contact when using the SWEAT manuals. Most soldiers stated that the manual checklists were bulky and difficult to write on in adverse weather conditions such as wind, rain, or bright sunlight. During our experiment, it was very bright and windy. Many soldiers had to position themselves in a way to prevent the pages from blowing away and seeking some shade to see the checklists. Additionally, soldiers felt that when they were in full battle gear (gloves, protective eye gear, weapon), it was very difficult for them to stop the assessment and assume a proper fighting position. We observed this throughout the experiment as well. Consequently, the GATER and HAMMER had very positive usability responses from soldiers. Both handheld devices are light, and easily fit into soldiers hands as seen in Fig. 5.

Soldiers could comfortably secure the device to themselves with a strap, which was not in the way of their weapon. During “react to contact,” soldiers could easily drop the device, draw their weapon, and assume a fighting position with the device still secured on their armament.

Fig. 5. Soldier in full battle gear with HAMMER

Recommend Device to Commander

Generally, there was nothing very different in the average responses of which tool type soldiers would recommend. We asked this question because we felt that a significantly different response may indicate that soldiers were much more enthusiastic about a particular tool. However, most of the answers were similar and not revealing about recommending any of the three tools. The most likely reason for this is that all of the tools were new systems that most of the soldiers were not familiar with. Furthermore, there were different features of each of the tools that the soldiers liked and disliked. Most agreed that a rugged digital device would be more ergonomic and probably increase overall efficiency. However, one of the attributes the soldiers did not particularly like with either the GATER or HAMMER was the poor quality images. For example, Fig. 6 represents an image of a pump data plate using GATER.

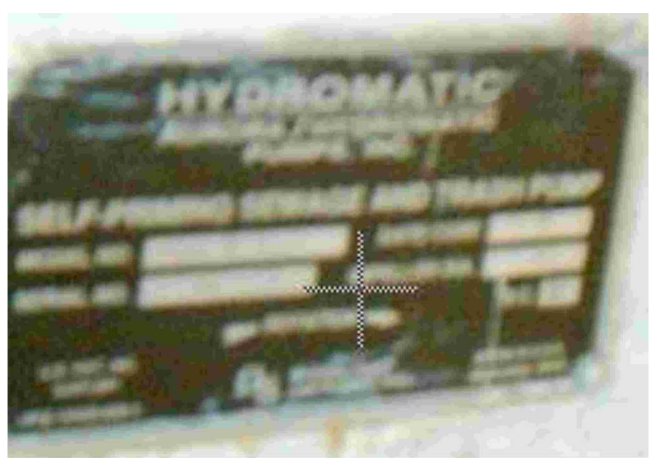


Fig. 6. Image with GATER

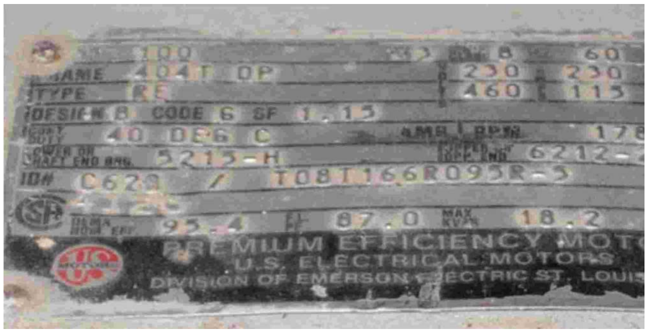


Fig. 7. Image with average digital camera

When the soldier took this image, he could instantaneously see the image on his handheld screen display. Because of the size and the average resolution on the screen display, most soldiers accepted the images and continued with their assessment. However, the actual image as downloaded on the PC for further analysis was poor in quality in many cases. The lack of a flash and the small number of megapixels on the integrated digital camera resulted in a cloudy picture as illustrated in Fig. 6. Had this been a "real world" mission, the BCT staff engineer would have to plan another mission to retrieve the data plate information, thereby, risking additional resources. By contrast, Fig. 7 is an image taken by a digital camera with a flash during a SWEAT IR manual evaluation.

Document More Using Particular Tool

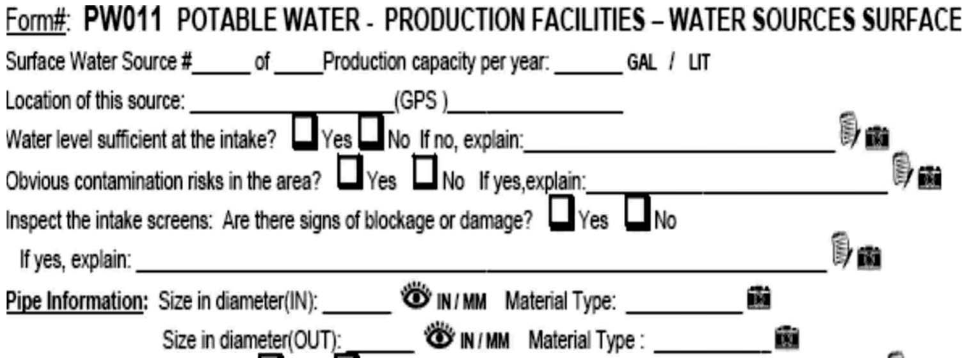
As mentioned previously, our second hypothesis was that soldiers would prefer the GATER because it was less restrictive and allowed them to take as much or as little information given the conditions on the ground. Most soldiers using the GATER commented that the system's broad data entry fields did not restrict the amount of information, and they could take multiple images. However, most soldiers agreed that the HAMMER enabled them to document more information than they otherwise would have if they used another tool followed by the SWEAT manual. There were a few reasons for this. First, most soldiers felt that a checklist-based assessment form helped them to look for specific items. Second, soldiers responded positively to the digital checklist because it helped guide them through the assessment, allowed them to quickly select entries using dropdown menus, required them to make certain entries, and allowed them to save the data per page unlike the soldiers using the GATER who had to type in all of the relevant information using the virtual keyboard.

Preference for Tool That Is More Difficult to Use but Provides Most Information

Another interesting finding that was different than what we originally expected was the overall opinion among most soldiers that a tool that was much more user friendly but did not yield the most information was not the most preferred. The survey results demonstrate different opinions among different tool users, most notably the group using the GATER. We feel that there are a few reasons for this difference in opinion. First, most of the more experienced combat engineers could see the value in a tool that facilitated rapid, yet more specific, data entry. Furthermore, most of the combat engineer officers felt that most of the difficulties with HAMMER were not necessarily in the actual data collection; rather, it was less user friendly from a staff officer perspective (programming the checklist with specific entries and subsequent

Table 7. Most Common Cognitive Demands—SWEAT

Task	Difficult cognitive element	Why difficult?	Common errors	Cues and strategies used
Initial setup	Recognizing the right form to use for a particular infrastructure site.	Soldiers are not very familiar with infrastructure facilities.	Starting with the wrong form.	Recognizing that most of the entries require information that is not recognizable.
Conducting assessment	Filling out all of the entries with appropriate information.	Entries are close together and difficult to read in bright sunlight or adverse weather.	Soldiers skip over entries and/or do not recognize icons.	Recognizing the important entries and skipping over the ones with the least impact on the results.
Reviewing assessment	Understanding handwriting or short hand symbols used to document data.	Soldiers rush through assessment.	Soldiers erase entries just because they are illegible.	Rewrite the assessment immediately after data collection.



Form#: **PW011** POTABLE WATER - PRODUCTION FACILITIES – WATER SOURCES SURFACE

Surface Water Source #_____ of ____Production capacity per year: ______ GAL / LIT

Location of this source: __________________(GPS)_______________

Water level sufficient at the intake? ☐ Yes ☐ No If no, explain:_______________________

Obvious contamination risks in the area? ☐ Yes ☐ No If yes,explain:_______________________

Inspect the intake screens: Are there signs of blockage or damage? ☐ Yes ☐ No

If yes, explain: ____

Pipe Information: Size in diameter(IN): ______ IN / MM Material Type: __________

Size in diameter(OUT):______ IN / MM Material Type : __________

Fig. 8. SWEAT IR form and icons

data processing). They saw little value in a system that provided soldiers with little instruction on what specifically to look for. The second reason why there was disparity in opinions within the group was that many of the participants were not combat engineers. For example, one noncommissioned officer who was an armor squad leader commented that the GATER was useful to him because he would probably use the GATER for other data collection missions other than infrastructure assessment. The flexibility of the GATER would better serve that purpose.

Further Insight into Usability

Another useful usability technique we used to further explain some of the users' difficult cognitive activities, why they were difficult, common errors associated with the different tasks, and the cues and strategies of different cognitive behaviors used to conduct the assessment, was in the cognitive demand tables (Militello and Hutton 1998). Of the 25 soldiers evaluated, we conducted nine cognitive demand tables. Table 7 is a summary of the most common cognitive demands of a typical small unit soldier using the SWEAT manual.

Cognitively, the SWEAT manual had the fewest demands that we would classify as very difficult. The SWEAT-IR manual was the only tool that had references and pictures that described certain components of an infrastructure facility. Most of the soldiers commented that this helped them to identify what components they were looking at in a facility. Additionally, the soldiers felt more confident about the manual since there were no technical issues that could prevent them from carrying out an assessment in a controlled environment. Last, most entries had icons that cognitively help soldiers take particular action such as measure, capture image, and document as seen in Fig. 8.

There were many cognitive demands associated with the HAMMER as seen in Table 8. However, we had to carefully delineate those that were more knowledge-based challenges rather than HCI related cognitive demands specific to the tool. Most of the cognitive demands had to do with the fact that the interface was not originally designed for the small unit soldier. Without proper training on the ArcPad system itself, users found it very difficult to know the proper sequencing of function keys if they had to troubleshoot the device or adjust the settings. For example, if the initial setting is such that the device requires multiple satellite coverage to take an image, users have to first understand why the system was preventing them from capturing data, and they had to know how to tab back to settings to select the proper adjustment. However, there were some positive cognitive observations of HAMMER users as well. The most commonly observed element was the ability of users to go through the facility and record data in a sequential manner. This was largely due to the fact that the HAMMER, unlike the SWEAT manual and GATER, had checklist "pages" that helped to guide the soldier through a facility he/she was relatively unfamiliar with. Furthermore, the dropdown menus significantly helped soldiers to quickly select entries using dropdown menus, required them to make certain entries, and allowed them to save the data easily.

Although both GATER and HAMMER applications run on the same hardware device (IKE 304), the process of turning the system on, getting to the right page, then starting the assessment seemed to flow much better for most soldiers and students. Cognitively, the soldiers felt that the features on the screen "forms" were much easier to understand than those using the HAMMER with customized forms. Table 9 is a summary of the most common demands using GATER.

One of the features that the soldiers could easily identify with was the black/red/amber/green color coded status entry, which was similar to other military status reporting methodologies. This feature was the only dropdown menu during the assessment itself. Some of the GATER users had the same cognitive challenges of knowing what functions to select to return to and adjust settings. However, the sequencing of those steps with GATER was designed, apparently, much more for the basic user. Another cognitive element that we observed with GATER users was how well they returned to their assessment after saving it to review for errors. With the aid of icons that represent points on a digital map on the display screen, users could "tap" the icon and review their assessment. Although many found this to be easy, others failed to carefully review to see if it was in fact their information saved or data from another assessment. This task was one that most users familiar with digital devices understood better because they were familiar with the basic zoom and pan features on the PDA.

Recommendations

Given the overall system of assessments within an engineering organization, we felt that a recommendation of specific tool would be too simplistic. Rather, we recommend the following with respect to training, data collection, and where each tool should be utilized within the BCT organization:

Training. We recommend at least three days of hands on training for each tool with realistic tasks, conditions, and standards to better understand the process and gain more confidence.

Table 8. Most Common Cognitive Demands—HAMMER

Task	Difficult cognitive element	Why difficult?	Common errors	Cues and strategies used
Initial setup	Set/verify startup settings; loading proper layers; understanding GPS functions.	Lack of training; top down menus complicated; use of stylus difficult.	Add the wrong layers; select wrong settings; many give up.	Prepare the device in advance; review the HELP menu.
Conducting assessment	Disengaging the PDOP when GPS not working.	Lack of satellites will prevent system from functioning unless PDOP turned off.	Taking to few images and skip over critical entries.	Verify images after capture; disengaging PDOP after initial GPS grid.
Reviewing assessment	Reviewing through several screens accurately.	If in a hurry, may skip over saved data.	Not identifying incorrect data efficiently.	Zooming in on PDA screen, carefully review data.

Table 9. Most Common Cognitive Demands—GATER

Task	Difficult cognitive element	Why difficult?	Common errors	cues and strategies used
Initial setup	Interface is confusing and loading time is lengthy.	Novice might not understand the interface or how to adjust settings.	Novice might attempt to capture data before application is loaded.	Starts the devices minutes before the assessment begins; conducts test assessment.
Conducting assessment	Taking correct images; taking correct distances using laser; properly describing overall condition.	If device is not aimed properly, images and measurements will be incorrect; soldiers not familiar with assessment descriptions.	Novice might not realize captured data is incorrect or incomplete; fail to describe overall condition of facility.	Recognize poor light conditions and adjust settings prior to collection; use down color coded condition feature.
Reviewing assessment	Retrieving one's assessment by selecting the icon on the PDA screen.	There may be several icons on the screen.	Novice may fail to zoom in on his/her assessment and recognize data.	Zoom in on the PDA screen, tab through the assessment methodically.

Data Collection. First, units should use a handheld device with a checklist for data collection. The device should have the following attributes and capabilities: (1) the screen should be slightly bigger to clearly see the data; (2) users should be able to easily override the GPS restriction once an initial image is captured; (3) schematics of basic infrastructure components should be hyperlinked to certain fields for quick reference; and (4) additional means of typing in data other than a virtual keyboard and stylus such as those found on cell phones with a dedicated keyboard.

Organization. Small unit leaders with infrastructure expertise and experience on new technologies should be primary data collectors for the BCT. Second, the GATER should be used for initial data collection, and subsequent missions requiring specific information should be collected with HAMMER. Third, all small unit leaders should have a pocket sized SWEAT manual for reference.

Conclusions

We conclude that handheld devices are useful for infrastructure assessment. Our findings are likely generalizable to broader populations than soldiers. While the conditions facing soldiers are unusual, they are not unique. In particular, there is likely a broad need for data collection by personnel with limited expertise, particularly in disaster conditions. Our finding of the advantages of handheld devices over paper and pencil methods suggests that further development of such devices is needed. Our research indicates that handheld devices allow for more data capture if the applications are designed to maximize the attributes of the device that differentiate it from paper-and-pen methods such as drop-down menus, multiple choice entries, and integration of GPS, digital imagery, and measuring technologies with text. While some of these attributes clearly reduce time, the technical complexities and system limitations (screen size and resolution, processor speed) can minimize some of the time saving advantages. A particularly useful finding is that there is a preference for more capable devices over those optimized for simplicity. This is not to say that good usability practices should be ignored, rather, there is a need for capable tools. At the same time, our findings do show that experience has value. The more experienced soldiers performed better on assessments than younger soldiers who, presumably, have more experience with handheld type devices. In all, our findings suggest the need for assessment tools that are developed to support rather than replace experience.

Acknowledgments

The writers thank the officers and soldiers of the 1st BCT, 1st CD, the 1-395 EN BN, the officers and cadets of West Point's Systems Department, CERL for their support on HAMMER, and EI2RC for their support with GATER.

References

Britt, T. (2006). "Utility assessment at Fort Jackson, South Carolina using the HAMMER." Construction Engineering Research Laboratory (CERL), Champagne, Ill.

Carayon, P., and Smith, M. J. (2000). "Work organization and ergonomics." *Appl. Ergon*, 31(6), 649–662.

Connelly, K. et al. (2005). "Designing a PDA interface for dialysis patients to monitor diet in their everyday life." *Proc., HCI Int.* (CD-ROM).

Converse, J. M., and Presser, S. (1986). *Survey questions: Handcrafting the standardized questionnaire*, Sage, Beverly Hills, Calif.

Karkkainen, L., and Laarni, J. (2002). "Designing for small screen display screens." *Proc., Nordic Conf. of Human Computer Interaction*, 227–230.

Militello, L. G., and Hutton, R. J. B. (1998). "Applied cognitive task analysis (ACTA): A practitioner's toolkit for understanding cognitive task demand." *Ergonomics*, 41(11), 1618–1641.

Newman, L. (2005). "Focus groups." *Handbook of human factors and ergonomic methods*, H. Stanton, A. Hedge, K. Brookhuis, E. Salas, and H. Hendrick, eds., CRC Press, Boca Raton, Fla., 78.1–78.5.

Pena-Mora, F., and Dwivedi, G. H. (2002). "Multiple device collaborative and real time analysis system for project management in civil engineering." *J. Comput. Civ. Eng.*, 15(4), 23–38.

Segall, N., Doolen, T. L., and Porter, J. D. (2004). "A usability comparison of PDA-based quizzes and paper and pencil quizzes." *Comput. and Educ.*, 45(4), 417–432.

Solomon, D. J. (2001). "Conducting web-based surveys." *Practical assessment, research and evaluation*, ⟨http://PAREonline.net/getvn.asp?v=7&n=19⟩.